Protein Folding

Protein folding is a process in which a polypeptide folds into a specific, stable, functional, three-dimensional structure. It is the process by which a protein structure assumes its functional shape or conformation Proteins are formed from long chains of amino acids; they exist in an array of different structures which often dictate their functions. Proteins follow energetically favorable pathways to form stable, orderly, structures; this is known as the proteins' native structure. Most proteins can only perform their various functions when they are folded. The proteins' folding pathway, or mechanism, is the typical sequence of structural changes the protein undergoes in order to reach its native structure. Protein folding takes place in a highly crowded, complex, molecular environment within the cell, and often requires the assistance of molecular chaperones, in order to avoid aggregation or mis folding. Proteins are comprised of amino acids with various types of side chains, which may be hydrophobic, hydrophilic, or electrically charged. The characteristics of these side chains affect what shape the protein will form because they will interact differently intra molecularly and with the surrounding environment, favoring certain conformations nd structures over others. Scientists believe that the instructions for folding a protein are encoded in the sequence. Researchers and scientists can easily determine the sequence of a protein, but have not cracked the code that governs folding.

Protein Folding theory and experiment

Early scientists who studied proteomics and its structure speculated that proteins had templates that resulted in their native conformations. This theory resulted in a search for how proteins fold to attain their complex structure. It is now well known that under physiological conditions, proteins normally spontaneously fold into their native conformations. As a result, a protein's primary structure is valuable since it determines the three-dimensional structure of a protein. Normally, most biological structures do not have the need for external templates to help with their formation and are thus called self-assembling.

Protein Renaturation

Protein renaturation known since the 1930s. However, it was not until 1957 when Christian Anfinsen performed an experiment on bovine pancreatic RNase A that protein renaturation was quantified. RNase A is a single chain protein consisting of 124 residues. In 8M urea solution of 2-mercaptoethanol, the RNase A is completely unfolded and has its four disulfide bonds cleaved through reduction. Through dialysisof urea and introducing the solution to O2 at pH 8, the enzymatically active protein is physically incapable of being recognized from RNase A. As a result, this experiment demonstrated that the protein spontaneously renatured.

One criteria for the renaturation of RNase A is for its four disulfide bonds to reform. The likelihood of one of the eight Cys residues from RNase A reforming a disulfide bond with its native residue compared to the other seven Cys residues is 1/7. Futhermore, the next one of remaining six Cys residues randomly forming the next disulfide bond is 1/5 and etc. As a result, the probability of RNase A reforming four native disulfide links at random is (1/7 * 1/5 * 1/3 * 1/1 = 1/105). The result of this probability demonstrates that forming the disulfide bonds from RNase A is not a random activity.

When RNase A is reoxidized utilizing 8M urea, allowing the disulfide bonds to reform when the polypeptide chain is a random coil, then RNase A will only be around 1 percent enzymatically active after urea is removed. However, by using 2-mercaptoethanol, the protein can be made fully active once again when disulfide bond interchange reactions occur and the protein is back to its native state. The native state of the RNase A is thermodynamically stable under physiological conditions, especially since a more stable protein that is more stable than that of the native state requires a larger activation barrier, and is kinetically inaccessible.By using the enzyme protein disulfide isomerase (PDI), the time it takes for randomized RNase A is minimized to about 2 minutes. This enzyme helps facilitate the disulfide interchange reactions. In order for PDI to be active, its two active site Cys residues needs to be in the -SH form. Furthermore, PDI helps with random cleavage and the reformation of the disulfide bonds of the protein as it attain thermodynamically favorable conformations.

Posttranslationally Modified Proteins Might Not Renature

Proteins in a "scrambled" state go through PDI to renature, and their native state does not utilize PDI because native proteins are in their stable conformations. However, proteins that are posttranslationally modified need the disulfide bonds to stabilize their rather unstable native form. One example of this is insulin, a polypeptide hormone. This 51 residue polypeptide has two disulfide bonds that is inactivated by PDI. The following link is an image showing insulin with its two disulfide bonds. Through observation of this phenomena, scientists were able to find that insulin is made from proinsulin, an 84-residue single chain. This link provides more information on the structure of proinsulin and its progression on becoming insulin. The disulfide bonds of proinsulin need to be intact before conversion of becoming insulin through proteolytic excision of its C chain which is an internal 33-residue segment. However according to two findings, the C chain is not what dictates the folding of the A and B chains, but instead holds them together to allow formation of the disulfide bonds. For one, with the right renaturing conditions in place, scrambled insulin can become its native form with a 30% yield. This yield can be increased if the A and B chains are cross-linked. Secondly, through analysis of sequences of proinsulin from many species, mutations are permitted at the C chain eight times more than if it were for A and B chains.

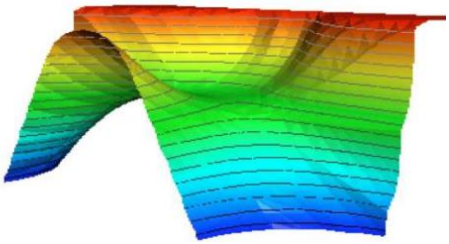
The Protein Folding Process

Considerable evidence suggests that all of the information to describe the three dimensional conformation of a protein is contained within the primary structure. However, for the most part, we cannot fully interpret the information contained within the sequence. To understand why this is true, we need to take a more careful look at proteins and how they fold.

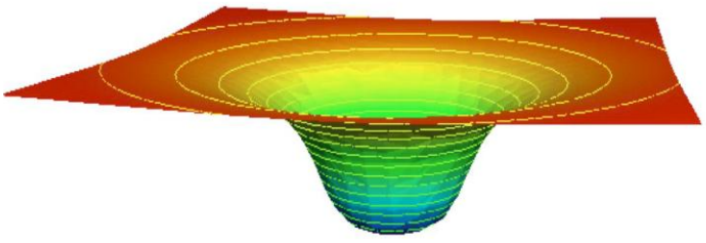
The polypeptide chain for most proteins is quite long. It therefore has ***many*** possible conformations. If you assume that all residues could have 2 possible combinations of ϕ and ψ angles (real peptides can have many more than this), a 100 amino acid peptide could have 2^{100} ($\sim 10^{30}$) possible conformations. If the polypeptide tested a billion conformations/second, it would still take over 10^{13} years to find the correct conformation. (Note that the universe is only $\sim 10^{10}$ years old, and that a 100 residue polypeptide is a relatively small protein.) The observation that proteins cannot fold by random tests of all possible conformations is referred to as the Levinthal paradox.

Folding pathways

In classical transition state theory, the reaction diagram for a spontaneous two state system is considered to have a high-energy starting material, a lower energy product, and an energy barrier between them. While the typical diagram that describes the process (such as the one shown at right) is useful, it is incomplete. The process for the conversion of S to P could actually take many pathways; the pathway shown is merely the minimum energy route from one state to another. The true situation is described by an energy landscape, with the minimum energy route being the equivalent of a pass between two mountains. Thus, although the pathway involves an energy barrier, other pathways require passing through even higher energy states.



A large part of the reason that single pathways (or small numbers of pathways) exist for chemical reactions is that most reactions involve the cleavage and reformation of covalent bonds. The energy barrier for breaking a covalent bond is usually quite high. In protein folding, however, the interactions involved are weak. Because the thermal energy of a protein molecule is comparable to the typical non covalent interaction strength, an unfolded polypeptide is present in a large variety of rapidly changing conformations. This realization led to the Levinthal paradox: because the unfolded protein should be constantly changing its shape due to thermal motions of the different parts of the polypeptide, it seemed unlikely that the protein would be able to find the correct state to begin transiting a fixed folding pathway.



An alternate hypothesis has been proposed, in which *portions* of the protein self-organize, followed by folding into the final structure. Because the different parts of the protein begin the folding process independently, the shape of the partially folded protein can be very variable. In this model, the protein folds by a variety of different paths on an energy landscape. The folding energy landscape has the general shape of a **funnel**. In the folding process, as long as the overall process results in progressively lower energies, there can be a large variety of different pathways to the final folded state.

The folding funnel shown above has a smooth surface. Actual folding funnels may be fairly smooth, or may have irregularities in the surface that can act to trap the polypeptide chain in misfolded states. Alternatively, the folding funnel may direct the polypeptide into a ***metastable*** state. Metastable states are local minima in the landscape; if the energy barriers that surround the state are high enough, the metastable state may exist for a long time – metastable states are stable for **kinetic** rather than **thermodynamic** reasons.

The difficulty in refolding many proteins *in vitro* suggests that the folded state of at least some complex proteins may be in a metastable state rather than a global energy minimum.

Folding process

The lower energies observed toward the depression in the folding funnel are thought to be largely due to the collapse of an extended polypeptide due to the hydrophobic effect. In addition to the hydrophobic effect, de solvation of the backbone is necessary for protein folding, at least for portions of the backbone that will become buried. One method for desolvation of the backbone is the formation of secondary structure. This is especially true for helical structures, which can form tightly organized regions of hydrogen bonding while excluding water from the backbone structure. A general outline for the process experienced by a folding protein seems to look like this:

A general outline for the process experienced by a folding protein seems to look like this:

1. Some segments of a polypeptide may rapidly attain a relatively stable, organized structure (largely due to organization of secondary structural Elements).
2. These structures provide nuclei for further folding.
3. During the folding process, the protein is proposed to form a state called a **Molten globule**. This state readily rearranges to allow interactions between different parts of the protein.
4. These nucleated, partially folded domains then coalesce into the folded protein. If this general pathway is correct, it seems likely that at least some of the residues within the sequence of most proteins function to guide the protein into the proper folding pathway, and prevent the “trapping” of the polypeptide in unproductive

Partially folded states.

Folding inside cells

Real cells contain **many** proteins at a high overall protein concentration. The protein concentration inside a cell is ~150 mg/ml. folding inside cells differs from most experiments used to study folding *in vitro*:

1. Proteins are synthesized on ribosomes. The entire chain is not available to fold at once, as is the case for an experimentally unfolded protein in a test tube.

2. Within cells, the optimum ionic concentration, pH, and macromolecule Concentration for each protein to fold properly cannot be controlled as tightly as in an experimental system.

3. Major problems could arise if unfolded or partially folded proteins encountered one another. Exposed hydrophobic regions might interact, and form potentially lethal insoluble aggregates within the cell.

One mechanism for limiting problems with folding proteins inside cells volves specialized proteins called **molecular chaperones**, which assist in folding proteins. Molecular chaperones were first observed to be involved in responses to elevated temperature (*i.e*. "heat shock") to stabilize existing proteins and prevent protein aggregation and were called heat-shock proteins (abbreviated as "hsp"). Additional research revealed that heat shock proteins are present in all cells, and that they decrease or prevent non-specific protein aggregation and assist in protein folding.

Thermodynamics of protein folding

In contemplating protein folding, it is necessary to consider different types of amino acid side-chains separately. For each situation, the reaction involved will be assumed to be:

Protein *unfolded* ⟷ Protein *folded*

Note that this formalism means that a negative ΔG implies that the folding process is spontaneous.

First we will look at **polar groups** in an aqueous solvent. For polar groups, the ΔH_{chain} favors the unfolded structure because the backbone and polar groups interact form stronger interactions with water than with themselves. More hydrogen bonds and electrostatic interactions can be formed in unfolded state than in the folded state. This is true because many hydrogen bonding groups can form more than a single hydrogen bond. These groups form multiple hydrogen bonds if exposed to water, but frequently can form only single hydrogen bonds in the folded structure of a protein.

For similar reasons, the $\Delta H_{solvent}$ favors the folded protein because water interacts more strongly with itself than with the polar groups in the protein. More hydrogen bonds can form in the absence of an extended protein, and therefore the number of the **sum of the ΔH polar** contributions is close to zero, but usually favors the folded structure for the protein slightly.

The chain ΔH contributions are positive, while the solvent ΔH contributions are negative. The sum is slightly negative in most cases, and therefore slightly favors folding.

The Δ*S chain* of the polar groups favors the unfolded state, because the chain is much more disordered in the unfolded state. In contrast, the Δ*S solvent* favors the folded State, because the solvent is more disordered with the protein in the folded state. In most cases, the **sum of the ΔS polar** favors the **unfolded** state slightly. In other words, the ordering of the chain during the folding process outweighs the other entropic factors.

The **ΔG polar** that is obtained from the values of ΔH polar and ΔS polar for the polar groups varies somewhat, but usually tends to favor the unfolded protein. In other words, the folding of proteins comprised of polar residues is usually a nonspontaneous process.

Next, we will consider a chain constructed from **non-polar groups** in aqueous Solvent. Once again, the Δ*H chain* usually favors the unfolded state slightly. Once again, the reason is that the backbone can interact with water in the unfolded state. However, the effect is smaller for non-polar groups, due to the greater number of favorable van der Waals interactions in the folded state. This is a result of the fact that non-polar atoms form better van der Waals contacts with other non-polar groups than with water; in some cases, these effects mean that the Δ*H chain* for nonpolar residues is slightly negative.

As with the polar groups, the Δ*H solvent* for non-polar groups favors the folded state. In the case of non-polar residues, Δ*H solvent* favors folding more than it does for polar groups, because water interacts much more strongly with itself than it does with non-polar groups,

The **sum of the ΔH non-polar** favors folding somewhat. The magnitude of the ΔH nonpolar is not very large, but is larger than the magnitude of the ΔH polar, which also tends to slightly favor folding.

The Δ*S chain* of the non-polar groups favors the less ordered unfolded state. However, the Δ*S solvent* highly favors the folded state, due to the hydrophobic effect. During the burying of the non-polar side chains, the solvent becomes more disordered. The ΔS *solvent* is a major driving force for protein folding.

The **ΔG non-polar** is therefore negative, due largely to the powerful contribution of the Δ*S solvent*.

Adding together the terms for ΔG polar and ΔG non-polar gives a slightly negative overall ΔG for protein folding, and therefore, proteins generally fold spontaneously raising the temperature, however, tends to greatly increase the magnitude of the TΔ*S chain* term, and therefore to result in unfolding of the protein.

The folded state is the sum of many interactions. Some favor folding, and some Favor the unfolded state. The qualitative discussion above did not include the magnitudes of the effects. For real proteins, the various ΔH and ΔS values are difficult to measure accurately. However, for many proteins it is possible to estimate the overall ΔG of folding. Measurements of this value have shown that **the overall ΔG for protein folding is very**

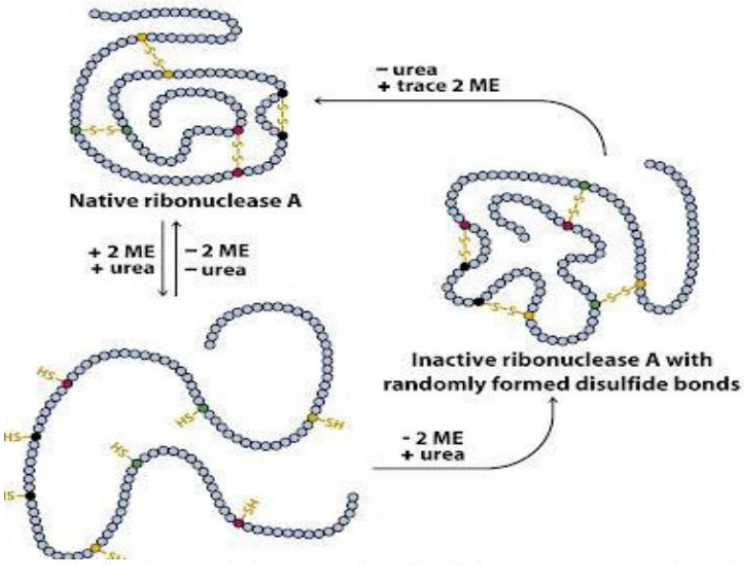
small: only about –10 to –50 k Joules/mol. This corresponds to a few salt bridges or hydrogen bonds.

Studies of protein folding have revealed one other important point: the hydrophobic effect is very important, but it is relatively non-specific. Any hydrophobic group will interact with essentially any other hydrophobic group. While the hydrophobic effect is a major driving force for protein folding, it is the constrains imposed by the more geometrically specific hydrogen bonding and electrostatic interactions in conjunction with the hydrophobic interactions that largely determine the overall folded structure of the protein.

Anfinsen's experiment

1. The Observation

Ribonuclease A (RNaseA) is an extracellular enzyme of 124 residues with four disulfide bonds. In the first phase of the experiment, the S-S bonds were reduced to eight –SH groups (using mercaptoethanol, HS-CH2-CH2-OH); the protein was then denatured with 8 M urea. Under these conditions, the enzyme is inactive and becomes a flexible random polymer. In the second phase, the urea was slowly removed (dialysis); then the –SH groups were oxidized back to S-S bonds. If the protein was able to regain its native structure spontaneously after removal of the urea, we expect that it would also regain its activity. In fact, the activity was >90% of the untreated enzyme. Moreover, sequence analysis showed that nearly all of the correct S-S bonds had been formed.



2. The Control

A reasonable objection can be raised to the above result by suggesting that perhaps RNase A was not completely unfolded in 8 M urea. To address this class of objections, RNase A was first reduced and denatured as above. But in the second phase, the enzyme was first oxidized to form S-S bonds, and then the urea was removed, i.e. the order of steps in the second phase of the experiment was reversed. The resulting activity was only about 1-2% of the untreated enzyme. Sequence analysis showed a random assortment of S-S bonds.

Anfinsen's work showed convincingly that proteins can indeed adopt their native information spontaneously, i.e. sequence determines structure. His demonstration of this fundamental property of proteins opened the problem to a massive amount of experimental and theoretical effort

The Levinthal paradox and kinetics

Levinthal's paradox is a thought experiment, also constituting a self-reference in the theory of protein folding. In 1969, Cyrus Levinthal noted that, because of the very large number of degrees of freedom in an unfolded polypeptide chain, the molecule has an astronomical number of possible conformations. An estimate of 3^{300} or 10^{143} was made in one of his papers. For example, a polypeptide of 100 residues will have 99 peptide bonds, and therefore 198 different phi and psibond angles. If each of these bond angles can be in one of three stable conformations, the protein may misfold into a maximum of 3^{198} different conformations (including any possible folding redundancy). Therefore if a protein were to attain its correctly folded configuration by sequentially sampling all the possible conformations, it would require a time longer than the age of the universe to arrive at its correct native conformation. This is true even if conformations are sampled at rapid (nanosecond or picosecond) rates. The "paradox" is that most small proteins fold spontaneously on a millisecond or even microsecond time scale. This paradox is central to computational approaches to protein structure prediction.

Levinthal himself was aware that proteins fold spontaneously and on short timescales. He suggested that the paradox can be resolved if "protein folding is sped up and guided by the rapid formation of local interactions which then determine the further folding of the peptide; this suggests local amino acid sequences which form stable interactions and serve as nucleation points in the folding process."[4] Indeed, the protein folding intermediates and the partially folded transition states were experimentally detected, which explains the fast protein folding. This is also described as protein folding directed within funnel-like energy landscapes some computational approaches to protein structure prediction have sought to identify and simulate the mechanism of protein folding. Levinthal also suggested that the native structure might have a higher energy, if the lowest energy was not kinetically accessible. An analogy is a rock tumbling down a hillside that lodges in a gully rather than reaching the base.

Protein Folding Rate

Determining how a protein will fold has been fairly difficult to predict even though the amino acid sequence is known. Instead of analyzing the structure of the protein and analyzing the mechanism of how a protein folds, understanding the kinetics of folding rates has proven to be a much more efficient way of understanding protein folding. The two-state folding kinetics of proteins is mostly studied, which analyzes the folding progress of a protein from its linear chain form, its primary structure, to its folded state, its tertiary structure. This process is dependent on the cooperative nature of the transition state. The kinetics of protein folding can be illustrated through the funnel energy landscape diagram, which is mathematically explained through the Gibbs free energy equation. This energy landscape diagram can follow the tract of the many pathways a protein can take until it reaches its

native, or most stable, folded state. As a protein conforms to its most native state, a free energy barrier ends up controlling the kinetics of the protein folding. To illustrate the folding mechanisms, different Go-model simulations are used, which are coarse-grained topology-based models. However, although Go-model simulations provide the folding mechanism of proteins, they lack the ability to predict the folding rates of proteins based on the kinetic or thermodynamic cooperativity demonstrated by two-state proteins. Because of this reason, studies have been done to understand the cooperative nature of the two-state folding of proteins and the factors that affect the folding rates of proteins.

Folding rate trends of Protein

The folding rates of two-state proteins can be understood through two general properties of the folded conformations. One of the trends is that more structurally complex proteins tend to fold at slower rates in comparison to more simple structural proteins. For example, a tertiary structure containing beta sheet proteins and proteins combined with alpha helices and beta sheets tend to fold slower than proteins that are made up of only alpha helices. The second trend is that larger proteins tend to fold a lot more slowly than smaller proteins. The kinetics of alpha helical proteins and structurally complicated proteins such as globular proteins also differ due to long-range tertiary contacts. The transition states of globular proteins are expected to have a higher transitional energy barrier than alpha helical proteins because more entropic energy is required to make a more structurally complicated protein to fold in a more ordered fashion in comparison to a simpler structural protein. As the chain length of a protein also increases, the free energy barrier exponentially increases as well to reach the transition state of the protein.

In determining the transition state of an in-process folded protein, the native state topology of the protein has to be known in order to predict the structure of the transition state of the protein. Topology refers to the effect of the orientation of objects in space due to deformations of the objects. In the case for proteins, a folded structure might change its orientation in space if the protein is heated up as it would lead to denaturing. To examine this transition state of folded proteins, the formation of the transition state is determined by the free energy barrier that controls the kinetics of the folding reaction. This free energy barrier is the result of the compensation of energy and the loss in entropy due to the new interactions formed in the process of protein folding. The relationship between the kinetics of a folding protein and topology help to explain why the transition state of a protein is dependent upon its native state. This is known as the principle of minimum frustration of energy landscape theory, which can related to the funnel model of folded proteins. The more stable the protein is, the lower the energy it is at, and the energy of the native protein can help give information on how much energy is required for a protein to reach its transition state in the folding process.

COOPERATIVITY OF PROTEINS

The use of Go models helps to give an identification of a protein in its most native state, which is held together by stabilizing interactions between native contacts. These stabilizing interactions are also known as non-additive forces, and these forces play a factor in the kinetics and thermodynamics of protein folding. These non-additive forces can also be thought of as intramolecular interactions that happen spontaneously within the protein such as side-chain ordering and hydrophobic forces. The effect of these non-additive forces have been shown to increase the free energy barrier of the two-state folded protein, and therefore, this makes these Go models more thermodynamically cooperative.

Upon using these Go models, the three-body interactions of the folding rates and what are known as phi values are examined in two-state proteins. The meaning of these phi values gives a relationship between the transition state of a two-state folded protein and its native state. The phi value explains the content of the native structure in its transition state. Therefore, the more native-like the structure of the transition state, the more likely this transition state will conform into its native state in a shorter period of time. In general, phi values improve when the transition state is more like its native state, but the ratio between its transition state and native state is different for each protein that varies in size and its secondary structure.

Many different types of Go models have been developed to better understand the cooperativity of the folding rates of proteins. For example, a Go model has been created in analyzing a small alpha-helical protein also known as a Calpha Go-like model. This model has also been altered by introducing solvent-mediated interactions to the model. The interactions between proteins are instead replaced by a desolvation barrier. Studies have shown that the thermodynamic and kinetic cooperativity of two-state folded proteins increase as the desolvation barrier increases in height. Desolvation is known as the removal of solvent from a material in solution. In general, desolvation has a property whee short-range contact proteins such as those that form alpha-helices have little cooperativity due to desolvation while long-range contacts such as those with a mix of beta sheets and alpha helices are expected to have high cooperativity because long-range contacts require persistence in bringing the proper chains together, and therefore, require a high amount of cooperativity. In conclusion, it is these topological models with nonadditive forces such as hydrophobic forces of proteins that help to better understand the folding rates of certain proteins.

MOLTEN GLOBULE

The molten globule state is an intermediate conformational state between the native and the fully unfolded states of a globular protein . Many proteins can be observed in this state when partially unfolded at equilibrium, under mild denaturation conditions, or as a transient intermediate kinetic species, being formed rapidly from the unfolded state upon transfer to refolding conditions. The characteristics of the molten globule state are:

1. the presence of a native-like content of secondary structure

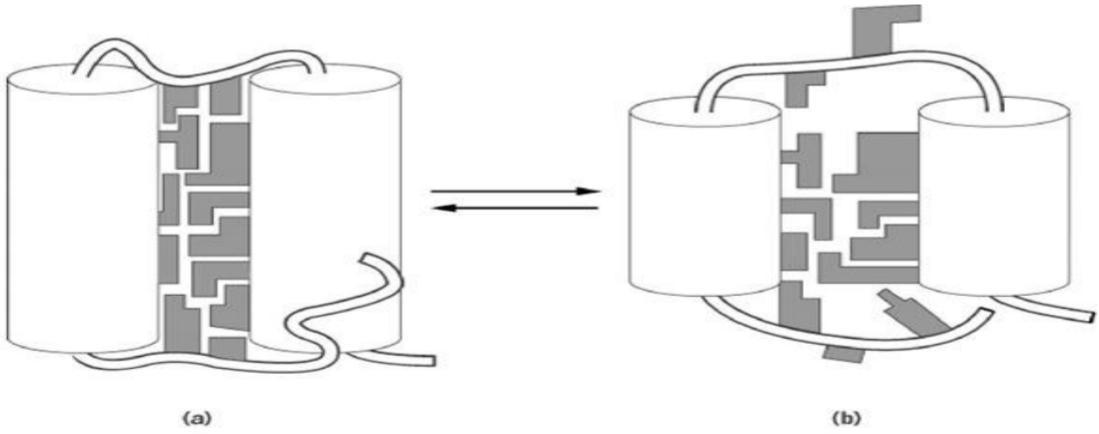
2. the absence of a specific tertiary structure produced by the tight packing of amino acid side chains

3. compactness in the overall shape of the protein molecule, with a radius 10 to 30% larger than that of the native state

4. the presence of a loosely packed hydrophobic core that increases the hydrophobic surface area accessible to solvent.

Thus, in short, the molten globule is a compact globule with a "molten" side-chain structure that is primarily stabilized by nonspecific hydrophobic interaction

Experimentally, the molten globule state is characterized by having a native-like circular dichroism (CD) peptide spectrum below 250 nm, which arises from the secondary structure, and by an unfolded-like CD spectrum of the aromatic side chains between 250 and 320 nm, due to the absence of specific side-chain packing interactions. Hydrogen atoms of the peptide backbone involved in secondary structure of the molten globule state appear to be protected from hydrogen exchange with the solvent protons, but the protection factor for the molten globule (10 to 1000) is much smaller than that for the native state, which is often greater than 10 6 The nuclear magnetic resonance (NMR) spectrum of the molten globule state is closer to that of the unfolded protein, and there is little, if any, chemical shift dispersion in the spectrum, reflecting the absence of a specific tertiary structure. The individual resonances in the NMR spectrum are, however, broader than those in the unfolded state,



A schematic model of the native (a) and the molten globule (b) states of a protein molecule.

reflecting conformational fluctuations in the molten globule state. It is known that the structure of the molten globule state, as determined by hydrogen exchange and NMR spectra, is heterogeneous in proteins, including a-lactalbumin, cytochrome c, and apo myoglobin. In these proteins, one portion of the structure is more organized and substantially protected in

the molten globule state, with other portions of the structure being less organized. Solution X-ray scattering has been used to characterize the molten globule structure. The presence of a clear peak in the Kratky plot and the radius of gyration evaluated from the Guinier plot of the X-ray scattering curve in the molten globule state show that the protein molecule in this state is compact and globular. Limited proteolysis by proteolytic enzymes has also been used for probing the partly folded structures of proteins; the key result is that the molten globule can be sufficiently rigid to prevent extensive proteolysis and it appears to maintain significant native-like structure. A hydrophobic fluorescent dye, such as 8-anilinonaphthalene-1-sulfonate (ANS), which binds to solvent-accessible hydrophobic surfaces of a protein molecule, has also been used for characterizing the molten globule state. The molten globule binds ANS much more strongly than does either the fully folded or fully unfolded form of the protein; the latter can be generated in a concentrated solution of a strong denaturant (6 M guanidinium chloride or 8 M urea).

For some proteins, the molten globule state is an equilibrium intermediate observed at an intermediate concentration of a strong denaturant (eg, 2 M guanidinium chloride) as part of a denaturant-induced unfolding transition. On the other hand, many globular proteins show a cooperative two-state unfolding transition without the intermediate. Whether or not the molten globule state is observed as a stable intermediate depends upon its stability relative to that of the native and the unfolded states. The unfolding intermediates of carbonic anhydrase and of a-lactalbumin are typical examples of a molten globule state that is stably populated at an intermediate concentration of denaturant. For these proteins, the partially unfolded states at acidic or alkaline pH are identical to the unfolding intermediate in the denaturant-induced transition, so the acidic or alkaline transitions also produce the molten globule state. For some other proteins, such as cytochrome c, apomyoglobin, and b-lactamase, the acidic or alkaline transition is known to produce a more extensively unfolded state; in these cases the addition of salt refolds the protein molecule from the unfolded to the molten globule state. The salt-induced refolding to the molten globule state is caused by counterion binding of the salt to the protein molecule, which eliminates the electrostatic repulsion between the charged groups. Other mildly denaturing processes that lead to the molten globule state include denaturation induced by hydrostatic pressure and by alcohols.

Removal of the bound metal ion in a metal-ion binding protein sometimes results in a molten globule state, as in the case of apo-a-lactalbumin produced by removal of the bound Ca. Covalent modification of a protein can also sometimes result in a molten globule state.

In many globular proteins, the molten globule state is observed at an early stage of the kinetics of refolding from the unfolded state. The early formation of the molten globule might

be a way by which these proteins can be folded efficiently without wandering into the huge conformational space available for the proteins. Two experimental techniques, stopped-flow CD and pulsed hydrogen-exchange combined with either two-dimensional NMR or electrospray ionization mass spectrometry, have been used successfully to characterize the transiently formed molten globule-like states during the kinetics of refolding of many globular proteins. The stopped-flow CD studies have shown the rapid formation of the peptide secondary structure occurring within the dead-time of the stopped-flow mixing (~10ms), although how much of the secondary structure is rapidly regained depends on the protein species. The pulsed hydrogen-exchange technique, when combined with two-dimensional NMR, can identify the specific location of stabilized secondary structure segments in a transient intermediate of a protein. For apomyoglobin and ribonuclease HI, comparison of the kinetic refolding intermediate and the equilibrium molten globule state has shown that the two are identical. Identification of the molten globule intermediate has also been well-established for a-lactalbumin by time-resolved CD and NMR studies. Molten globule-like folding intermediates have also been detected and characterized in many other globular proteins. Nevertheless, this does not necessarily mean that the molten globule state must be an obligatory, universal intermediate of protein folding. Because the formation of the molten globule-like folding intermediate is usually too rapid to be coupled kinetically with the subsequent folding reactions, it is very difficult to determine whether or not the molten globule is an obligatory folding intermediate. Furthermore, several small globular proteins with approximately 60 amino acid residues are known to refold very rapidly to the native state within a few milliseconds without accumulation of the molten globule intermediate.

When considering the role of the molten globule state in protein folding, it is important to address the question as to whether or not the molten globule is a thermodynamic state. Analysis of the cooperativity parameters for denaturant-induced unfolding transitions of some proteins has suggested that the transitions from the molten globule state to the unfolded state and from the native state to the molten globule state are both all-or-none transitions, indicating that the molten globule state is a thermodynamic state. Furthermore, stability studies of mutants of apomyoglobin and cytochrome c have concluded that the molten globule states of these proteins show cooperative unfolding and are stabilized by native-like tertiary interactions, in addition to nonspecific hydrophobic interactions, which is consistent with the proposal that the molten globule state is a distinct thermodynamic state. However, for the best-characterized molten globule of a-lactalbumin, calorimetry, NMR, vibrational Raman spectroscopy, and other techniques have clearly shown that the unfolding of this molten globule is not a cooperative two-state transition. Such diversity in the unfolding behavior of the molten globule state among different proteins may arise from the diversity of the molten globule structure. Because the native tertiary interactions are at least partially lost in the molten globule state, its structure must be more diverse than the native structure, and

how cooperatively the molten globule unfolds may depend on how many residual native tertiary interactions are retained in this state. A good example is the molten globule state of the equine lysozyme, which is a calcium-binding protein and homologous to a-lactalbumin. Although the equine lysozyme molten globule apparently resembles that of a-lactalbumin, a rigorous analysis of its spectroscopic and thermodynamic properties has shown that its structure is significantly more highly organized and that its unfolding is a cooperative first-order transition accompanied by a large change in enthalpy. Because of this diversity in the intermediate conformational states of proteins, it is difficult to provide a clear structural definition of the molten globule state. Consequently, this causes some controversy, with the formation of native-like tertiary fold considered to be a characteristic of the molten globule state in certain cases, while in other cases structures with non-native tertiary folds are also molten globules. Furthermore, more than one intermediate conformational state is often observed between the native and fully unfolded states. Nevertheless, the four characteristics itemized at the beginning of this article are those generally accepted as the characteristics of the molten globule state.

It is now well established that not only the native state, but also non-native conformational states, play an important role in a biological cell. The protein states recognized by various molecular chaperones are non-native. The non-native conformation is also required for translocation of a protein across a biological membrane. Various genetic diseases can be caused by the misfolding of translated polypeptides, and this misfolding results from an increased propensity of the mutant proteins to form non-native conformations. Because the molten globule state is regarded as a denatured state under physiological conditions, it definitely assumes some role in the above phenomena in vivo. It is also true, however, that there is a much greater diversity in the non-native conformations of proteins than in just the conformations characterized as the molten globule state.

BIOPHYSICAL TECHNIQUES FOR THE STUDY OF PROTEIN FOLDING

Rapid Mixing Methods

The mechanism of protein folding can be studied by two different groups of approaches. Equilibrium methods provide information about possible folding intermediate states or deduce rate constants from the molecular fluctuations or dynamic properties of the system. Relaxation methods follow the change of the system evolving toward a new equilibrium after a rapid perturbation of its extrinsic variables, such as temperature, pH, pressure, or solvent composition. The time required for folding varies greatly among proteins, ranging from microseconds to minutes. The smallest protein molecules with no folding

intermediates fold on the microsecond timescale, which, on one hand, might make them suitable for in silico-folding simulation studies. On the other hand, such fast reactions make the experimental detection of events during the folding process difficult. Basic techniques in the study of folding kinetics are stopped-flow fluorescence spectroscopy and stopped-flow circular dichroism. These techniques are capable of monitoring the formation of secondary and tertiary structures during the folding reaction with millisecond time resolution. In such experiments, rapid processes occurring within the dead time of the measurement were observed for many proteins. The challenge to resolve this initial burst phase and to reveal the structural changes taking place during the first millisecond stimulated the development of new, rapid kinetic techniques capable of triggering and monitoring the folding process on the sub- millisecond timescale. While the conventional stopped-flow apparatus, in which a small volume of a freshly made mixture containing the reacting components is injected into the measurement cell, is quite economical and offers a wide range of applications, its dead time is usually about 1 millisecond or longer. Continuous-flow methods extend the time resolution to the microsecond time range. In the continuous-flow cell, solutions are mixed under highly turbulent conditions to achieve complete mixing. The kinetics of the reaction is monitored under steady-state flow conditions as a function of the distance downstream from the mixer by using relatively simple and inexpensive detection methods. Using this technique, it has become possible to study the initial collapse and formation of intermediates in the early stage of the folding reaction, during the burst phase.

Real-Time NMR Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy has greatly contributed to our understanding of the protein-folding problem. Hydrogen–deuterium exchange experiments, revealing dynamical events at an atomic level, have illuminated the process of unfolding from the native state and the structure of folding intermediates. The quenched-flow pulse- labeling technique has enabled researchers to study the early stages of protein folding using a conventional NMR instrument. NMR studies have characterized the properties of denaturant-induced equilibrium folding intermediate states such as the molten globule. In equilibrium systems, the rates of conversion between distinct conformational states can be calculated from a line shape analysis of the NMR resonances, and therefore can provide kinetic data on folding. Slow folding reactions such as cis–trans prolyl isomerization can be directly followed by sequential recording of one-dimensional (1D) NMR spectra.This method is particularly useful for discovering intermediates formed at the late stages of the folding process. Using a stopped-flow device for injection of the protein solution into the NMR tube that already contains the denaturant or the refolding buffer pushes the dead time of mixing below 1 s . One of the first proteins studied by real-time NMR was a-lactalbumin. 1D-NOE (nuclear Overhauser effect) experiments revealed the native-like compactness of the transient molten globule state of a-lactalbumin. These experiments also demonstrated that the transient intermediate closely resembles the well-characterized stable molten globule state formed at

low pH. While 1D-NMR spectra have limited resolution, multidimensional NMR can provide high spatial resolution information on the folding process. Because recording multidimensional spectra is time consuming, only slow processes could be followed directly by sequential recording. Balbach and coworkers developed new methods to reconstruct the kinetic history of folding reactions from a single two-dimensional NMR spectrum recorded during the entire time course of the reaction. The basis of these methods is that the line widths and intensities reflect the history of the folding events occurring during spectral accumulation. When applied to a-lactalbumin, the technique demonstrated the cooperative nature of the folding of the main chain.

Reference

1. Wikipedia - https://en.wikipedia.org/wiki/Protein_folding
2. Protein Folding Handbook: Prof. Dr. Johannes Buchner and Prof. Dr. Thomas Kiefhaber First published:20 January 2005
3. Protein Folding First Edition by Thomas E. Creighton